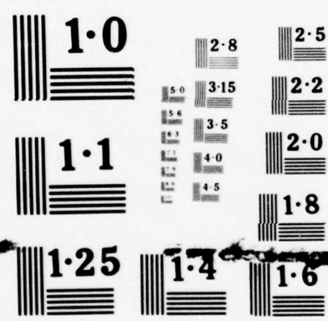
AD-A051 683 DECISIONS AND DESIGNS INC MCLEAN VA F/G 12/1
ON THE RECONCILIATION OF PROBABILITY ASSESSMENTS. (U)
FEB 78 D V LINDLEY, A TVERSKY, R V BROWN N00014-75-C-0426
UNCLASSIFIED TR-77-10-25 NL

1 OF 1
ADA
051683

END
DATE
FILMED
4 -78
DDC



NATIONAL BUREAU OF STANDARDS
MICROCOPY RESOLUTION TEST CHART

B.S.

February 1978

TECHNICAL REPORT TR 77-10-25

On the Reconciliation of Probability Assessments

D.V. Lindley
A. Tversky
R.V. Brown

DECISIONS and DESIGNS, INC.

OFFICE OF NAVAL RESEARCH
CONTRACT NUMBER N00014-75-C-0426
CONTRACT AUTHORITY IDENTIFICATION
NR 197-029 (455)

White Section	X
Buff Section	☐
UNANNOUNCED	☐
JUSTIFICATION	
BY	
DISTRIBUTION/AVAILABILITY CODES	
AVAIL. and/or SPECIAL	
A	

2

9 TECHNICAL REPORT 14 TR-77-10-25

6

ON THE RECONCILIATION OF PROBABILITY ASSESSMENTS

by

10 Dennis V. Lindley, Amos Tversky and Rex V. Brown

Sponsored by
Office of Naval Research 15 N00014-75-C-0426

11 February 1978

12 62p.

DECISIONS and DESIGNS, INC.
Suite 600, 8400 Westpark Drive
McLean, Virginia 22101
(703) 821-2828

390 664

1B

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: TR 77-10-25	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): ON THE RECONCILIATION OF PROBABILITY ASSESSMENTS	5. TYPE OF REPORT & PERIOD COVERED: Technical
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Dennis V. Lindley, University College, London Amos Tversky, Hebrew University, Jerusalem Rex V. Brown, Decisions & Designs, Inc.	8. CONTRACT OR GRANT NUMBER(s): N00014-75-C-0426
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Decisions and Designs, Incorporated 8400 Westpark Drive McLean, Virginia 22101	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research 800 N. Quincy Street Arlington, VA 22217	12. REPORT DATE: February 1978 13. NUMBER OF PAGES: 63
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): UNCLASSIFIED 15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Decision Theory
Probability Assessment
Coherence Analysis

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper investigates the question of how to reconcile incoherent probability assessments, i.e., assessments that are inconsistent with the laws of probability. A general model for the analysis of probability assessments is introduced, and two approaches to the reconciliation problem are developed. In the internal approach, one estimates the subject's "true" probabilities on the basis of his assessments. In the external approach, an external observer updates his own coherent probabilities in the light of → next page

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

the assessments made by the subject. The two approaches are illustrated and discussed. Least-squares procedures for reconciliation are developed within the internal approach.

ABSTRACT

This paper investigates the question of how to reconcile incoherent probability assessments, i.e., assessments that are inconsistent with the laws of probability. A general model for the analysis of probability assessments is introduced, and two approaches to the reconciliation problem are developed. In the internal approach, one estimates the subject's "true" probabilities on the basis of his assessments. In the external approach, an external observer updates his own coherent probabilities in the light of the assessments made by the subject. The two approaches are illustrated and discussed. Least-squares procedures for reconciliation are developed within the internal approach.

CONTENTS

TABLES

ON THE RECONCILIATION OF PROBABILITY ASSESSMENTS

1.0 INTRODUCTION

Decisions often depend on the probability of uncertain events such as the result of an election, the state of the economy, the outbreak of war, the guilt of a defendant, or the outcome of a medical operation. Because such events are essentially unique, the assessment of their probability must be based on personal judgment. Thus, the human mind is used as an instrument for the assessment of uncertainty, much as the ruler, the pen balance, and the pendulum are used as instruments for the measurement of length, weight, and time. In fact, the axiomatic analysis of subjective probability (De Finetti, 1937; Ramsey, 1931; Savage, 1954) is closely related to the axiomatic analysis of the measurement of physical attributes like length, weight, and time.

To illustrate this relation, recall that the analysis of many physical attributes is based on some set of objects, ordered with respect to the attribute in question, and a physical operation of concatenation of objects. In the measurement of mass, for example, the ordering of objects is typically established by the use of a pen balance; and the concatenation operation is interpreted as the placing of two objects together in the same pen. The key axiom needed to establish an essentially unique additive scale of mass is

$$x > y \text{ iff } x \circ z > y \circ z$$

where > denotes the mass ordering and o denotes the concatenation of objects (see Krantz, Luce, Suppes & Tversky 1971).

Suppose that our object set consists of events, and > denotes the relation "more probable than" between events. The ordering of the events could be established directly by asking an individual which of two events he considers more probable. Alternatively, the ordering could be derived from preferences between gambles as in the Ramsey-Savage approach. If we interpret the operation 0 as the union of disjoint events, then the above condition becomes one of the critical axioms needed to ensure the existence of a subjective probability measure over the appropriate collection of events.

Despite the formal similarity between the processes, the measurement of subjective probability is considerably more problematic and less satisfactory than the measurement of physical attributes such as length or mass and psychological attributes such as loudness or brightness. Indeed, the assessment of subjective probability is beset with severe problems of both theoretical and practical nature. First, subjective probability is a measure of degree of belief, which reflects one's state of information. It is not only subjective but also variable since it can change from one situation to another. Second, it is not possible in general to obtain repeated independent measurements of subjective probability from the same individual because he is likely to remember his previous thoughts and responses. Consequently, there are no procedures for the measurement of belief that permit the application of the law of large numbers to reduce measurement errors.

The difficulties involved in applying standard measurement criteria of reliability and validity to the measurement of belief give rise to the questions of how to evaluate and improve assessments of subjective probability. Three types of criteria that could be called pragmatic, semantic and syntactic have been employed. Pragmatic tests refer to comparisons of assessments with reality, and they are

applicable whenever the assessed probability of an event (e.g., a royal flush in poker, or an accident on the highway) can be meaningfully compared to a value that is computed in accord with the probability calculus, or estimated from empirical data. Unfortunately, such tests cannot be applied in most cases of interest because of the difficulties involved in estimating so-called objective probabilities.

If pragmatic tests are not applicable, however, it is still possible to evaluate probability assessments in terms of a semantic criterion that pertains to the meaning of the probability scale. Clearly, there is no way of validating, for example, a meteorologist's single judgment that the probability of rain is 2/3. If the meteorologist is using the scale properly, however, we would expect that rain would occur on about 2/3 of the days to which he assigns a rain probability of 2/3. This criterion is called calibration. Formally, a person is calibrated if the proportion of correct statements, among those that were assigned the same probability, matches the stated probability, i.e., if his hit-rate matches his confidence. If only 1/2 of the days to which the meteorologist assigned a rain probability of 2/3 were rainy, then he is not calibrated. This does not mean that his assessments are worthless or non-informative; it merely indicates an improper use of the probability scale. In order to make use of his assessment, it should be remembered that his 2/3 actually means 1/2.

Besides the pragmatic and the semantic criteria, subjective probabilities should also obey syntactic rules; that is, the relations between assessments should be governed by the laws of probability. For example, if A and B are disjoint events, then the assessed probability of the event, A or B, should be equal to the sum of the assessed probabilities for A and for B. A set of probability assessments is (internally) coherent only if it is compatible with the

probability axioms. Coherence is clearly essential if we are to treat assessments as probabilities and manipulate them according to probabilistic laws.

Common observations and experimental studies (see Slovic, Fischhoff and Lichtenstein, 1977; Tversky and Kahneman, 1974) show that the pragmatic, semantic and syntactic criteria are not always satisfied. Thus, assessments of probability, produced by laymen and experts alike, are often inaccurate, uncalibrated and incoherent. But since subjective judgments constitute the major data base for the measurement of uncertainty, the question is not whether to accept subjective judgments at face value or reject them altogether, but rather how to debias and improve them. Procedures for the elicitation and debiasing of subjective probabilities have been discussed by several authors, among them Kahneman and Tversky (1978); Spetzler and Stael von Holstein (1975); and Winkler (1969). These procedures were designed to obtain probability assessments that are more accurate and better calibrated. The present paper is concerned with the problem of coherence, namely, how to reconcile probability assessments that are incoherent or mutually inconsistent. Before we formalize the problem, it is instructive to examine some examples.

<u>Example 1.</u> Consider the possible causes of death. Suppose H denotes heart failure, C denotes cancer, D denotes any other disease, and N denotes all causes of natural death. Let $\bar{N}$ denote the complement of N, etc. The following probability assessments, denoted by q, were actually made by one of the authors concerning the possible causes of death of another author:

$q(H) = .33,\ q(C) = .27,\ q(D) = .23\ q(\bar{N}) = .12.$

$q(H|N) = .41,\ q(C|N) = .31,\ q(D|N) = .28$

Note that these assessements are incoherent. First, $q(H)+q(C)+q(D)+q(\bar{N}) = .95$ instead of unity. Such failures of additivity are quite common when the number of events exceeds 2 or 3. Second, the conditional probability ratios do not coincide with the ratios of the corresponding unconditional probabilities. For example, $q(H|N)/q(C|N) = 1.32$, while $q(H)/q(C) = 1.22$. Clearly, the two ratios should be equal because both H and C are subsets of N. The assessor is now faced with the task of reconciling his assessments so as to achieve coherence. How should he do it? What additional information, if any, is required to reconcile the inconsistent assessments?

Example 2. Let C denote the occurrence of a major energy crisis in the United States during the next decade, and let q_1 and q_2 be two different assessments of the probability of C. These values may emerge from two different ways of thinking about the problem, say, one in terms of specific scenarios that could lead to an energy crisis, and one in terms of a particular economic model. Alternatively, q_1 and q_2 may represent the judgments of two experts about the probability of C. In general, of course, q_1 and q_2 do not coincide. Thus, we have to amalgamate the two estimates, that is, assess the probability of C in light of q_1 and q_2, denoted $p(C|q_1,q_2)$. Note that from a purely formal viewpoint, it is immaterial whether the two estimates were produced by one person

using two different methods, or by two different individuals. In both cases we need to reconcile the difference and produce a single estimate.

The problem of reconciling inconsistent observations is not unique to the measurement of belief. For example, a surveyor who uses a theodolite to measure distance faces a similar problem. Because of measurement errors, the assessments of angle and distance are generally incompatible with the laws of plane geometry. Hence, the surveyor must reconcile the inconsistent measurements to obtain a coherent set of estimates. His problem, however, is simpler because he can readily obtain repeated observations and thereby reduce errors of measurement. Although it is generally not possible to obtain independent repeated measurements of subjective probability, the analogy between the measurement of distance and the assessment of belief is instructive. In particular, it suggests the possibility of exploiting the constraints imposed by the probability laws to obtain improved estimates of subjective probability, much as the surveyor exploits the constraints imposed by plane geometry.

To illustrate this idea, suppose you have to assess, as in Example 2, the probability of a major energy crisis in the United States during the next decade, denoted C. There are several different approaches to the problem. You could adopt an intuitive, wholistic approach where you contemplate the energy situation, the United States economy, the international scene, etc., and make an intuitive estimate on the basis of these considerations. Alternatively, you might wish to develop an explicit model for the supply and demand of energy, in which p(C) can be expressed as a function of some parameters that either are known, or can be estimated. A third possibility, which lies somewhere between wholistic assessment and explicit modeling, is to decompose p(C) and assess the components separately. For example, let E denote

an oil embargo on the United States during the relevant time period. Following the decomposition approach, you could assess the probability of an energy crisis given an oil embargo $p(C|E)$, the probability of an energy crisis in the absence of an oil embargo $p(C|\bar{E})$, and the probability of an oil embargo $p(E)$. The overall probability of an energy crisis is given by

$$p(C) = p(C|E)p(E) + p(C|\bar{E})p(\bar{E}), \text{ with } p(\bar{E}) = 1-p(E).$$

Thus, the laws of probability allow us to compute $p(C)$ from $p(E)$ and the conditional probabilities $p(C|E)$ and $p(C|\bar{E})$, just as the laws of geometry allow us to compute the distance between A and B, say, from the distance between A and X, the distance between B and X, and the angle AXB. Just as the distance between A and B can be measured using a different auxiliary point from X, the probability of C can be computed using a different conditioning event from E. Thus, one may compute $p(C)$ from $p(C|D)$, $p(C|\bar{D})$ and $p(D)$, where D denotes the development of new effective methods for using solar energy.

The use of different conditioning events such as D and E to compute $p(C)$ can be viewed as two different ways of thinking about the target event C. A direct wholistic assessment of C represents a third way of looking at the problem. Generally, the different procedures yield different estimates of $p(C)$ that have to be reconciled. If each of the estimates, however, conveys some valid information that is not included in the others, then the precision associated with the reconciled value will be greater than that of theseparate estimates, in the same way that the precision in the location of a point increases by determining several different (inconsistent) bearings.

The present paper is concerned with the development of models for the reconciliation of incoherence (Brown & Lindley, 1977). We first outline a general framework for the analysis of probability assessment; we then investigate how it can be used to reconcile inconsistent judgments. We develop two approaches to the reconciliation problem, which we label internal and external. In the internal approach, the observed probability assessments are related to some internal coherent probabilities in a manner analogous to the relation between the observed score and the true score in test theory. Thus, it is assumed that the subject has, in some sense, a set of coherent probabilities that are distorted in the elicitation process. The internal approach is concerned, then, with the attempt to estimate the underlying "true" probabilities using the observed assessments. This approach also permits the calculation of the precisions associated with the reconciled values, given the precisions associated with the observed assessments.

In the external approach, we introduce an external entity, the investigator, who assesses his own coherent probabilities on the basis of the judgments produced by the subject. Here, the investigator plays a role that is similar to that of a surveyor who uses fallible measurements to produce a coherent set of distances.

In the next section we develop a general framework for the analysis of fallible probability assessments, and introduce the internal and the external approaches to the reconciliation problem. Section 3 illustrates the application of the basic model in two simple special cases. Least-squares procedures for the internal approach are discussed and illustrated in Section 4. The philosophical and practical problems associated with the present development are discussed in Section 5, along with directions for future research.

2.0 THE BASIC MODEL

2.1 Description of the Model

We are concerned with an individual or a subject, denoted S, who considers a sequence $A = (A_1, A_2, \ldots, A_m)$ of events about which he is uncertain. For example, consider a meteorologist contemplating the rain pattern for the next m days where A_i denotes rain on the i-th day.

We suppose that S wishes to describe his uncertainty about A through a coherent probability specification for A. Normative, or coherent, S therefore has a probability distribution $\pi(A)$ for A. It is important to notice that π obeys all the rules of the probability calculus. In the example, the meteorologist would assess the probabilities of all weather patterns over m days, e.g., $\pi(A_1A_2) = \pi(A_1)$ $\pi(A_2|A_1)$.

Real S is not necessarily coherent and his assessments of probabilities do not always obey the rules of the calculus; even if they do, they may be defective because of S's weakness as a probability appraiser. Thus, in the meteorological situation S may assess the probability of rain on day 1 as 0.4, on day 2 given that day 1 is wet as 0.8, and yet say that the probability of rain on both days is 0.2 and not 0.32 as implied by the above assessments and the demands of coherence. The assessed values will be described by a vector $q(A)$.

Our model therefore contains, in addition to S, three elements: A, $\pi(A)$ and $q(A)$. The first describes the world external to S, the second describes a coherent S and the third gives S's stated view of the world. In terms of the analogy with the measurement of length, q corresponds to the

observed measurements of distances and angles, and π could be regarded as the true distances between the points. Like their physical counterparts, q is directly observable but π is not.

In addition to the subject, S, we consider an investigator, N. Unlike S, N is coherent and his task is to reconcile S's stated values q and to provide an assessment of π. Alternatively, N could assess his own probabilities for A in the light of the information q provided by S. N can be thought of as the surveyor who uses Euclidean geometry to provide estimates of the true positions except that he uses the probability calculus instead of geometry. It is possible to interpret N either as part of S, or as external to S. In either case, N is a coherent observer of S.

From N's viewpoint, all the elements of our model, S, A, π and q are part of the external world about which he is uncertain and which are described by a probability distribution $p(A,\pi,q)$ of the uncertain quantities. (We prefer Schlaifer's term "uncertain quantity" to the more usual "random variable" because the expressions are fixed and not variable; but they are unknown and hence uncertain.) The notation $p(\cdot)$ will be reserved for N's probabilities: the Greek equivalent $\pi(\cdot)$ similarly refers to S's coherent probabilities; a different letter, q, is used for S's assessments which need not be coherent.

The joint distribution may conveniently be described in three stages. First there is the distribution of A, $p(A)$. Second there is the conditional distribution of π given A, $p(\pi|A)$; and third there is the conditional distribution of q given the other two elements, $p(q|\pi,A)$. These three distributions completely specify the joint distribution of the uncertain quantities and summarize the situation as far as N is concerned. Notice that each of the three distributions

corresponds to a different aspect of N's contemplation of the problem. His view of the world external to both N and S is described by $p(A)$; whereas S's, when coherent, is $\pi(A)$. His view of coherent-S as a probability appraiser is included in $p(\pi|A)$. Finally, $p(q|\pi,A)$ gives N's opinion of S as a measuring instrument when he is measuring his uncertainties about A. In the meteorological example, $p(A_1)$ is N's probability for rain on the first day; $p[\pi(A_1)|A_1]$ is N's probability that the meteorologist's true probability of rain on the first day is π when it truly will rain then. The final conditional distribution describes what N thinks the meteorologist will actually state when his true value is $\pi(A_1)$ and it will rain.

Since q differs only from π because of S's difficulties in articulating his probabilities and therefore describes S as a measuring instrument on π, it seems reasonable to suppose that the conditional distribution of q, given π, does not depend on the state of the world external to S described by A. That is, given π, q and A are independent, or

$$(1) \qquad p(q|\pi,A) = p(q|\pi).$$

Indeed, the major function of the unobservable "true" probability π is to stand between S and the external world so that the two are related only through π. This assumption is in the spirit of the standard measurement model, where different measurements of the same quantity are treated as independent--given the "true" measurement of the quantity.

In summary, our model involves

I. $p(A)$: N's appreciation of the world
II. $p(\pi|A)$: N's opinion of S as a probability appraiser
III. $p(q|\pi)$: N's opinion of S as a judge of his (S's) uncertainty about the world.

These core distributions are N's, and all the calculations with them are performed by N according to the rules of the probability calculus, that is, coherently. We now describe two approaches to the reconciliation problem, called the internal and the external approaches.

In the internal approach, the subject acts as his own investigator and attempts to estimate his "true" coherent probabilities π. In the external approach the investigator attempts to revise his own probabilities for A, based on the assessments made by the subject.

2.2 The Internal Approach

Here N is concerned only with π and q. We have[1]

(2) $$p(\pi) = \Sigma_A p(\pi|A)p(A)$$

and $p(q|\pi)$ directly, providing a complete probabilistic description of q and π.

By Bayes' Theorem we have

(3) $$p(\pi|q) \propto p(q|\pi)p(\pi)$$

as N's appraisal of coherent S after S has reported his assessments q.

This method can be used when S has made several probability assessments $q = (q_1, q_2, \ldots q_m)$, finds them to be incoherent--as in Example 1 or the meteorological problem above--and wishes to reconcile them to coherent values. The

---

[1]The abbreviated notation Σ_A here refers to a summation over a partition of the events in $\underline{A}$. Thus if $\underline{A} = (A_1, A_2)$ then the summation is over $A_1A_2, A_1\bar{A}_2, \bar{A}_1A_2$ and $\bar{A}_1\bar{A}_2$.

calculations leading to (3) enable him to do this. Furthermore, they provide a way of calculating the precisions of the reconciled values by means of the variances of $p(\pi|q)$.

Let $(\pi_1, \pi_2, \ldots. \pi_m)$ be the true values that correspond to $(q_1, q_2, \ldots, q_m)$, respectively. There will typically be constraints among the π_i's corresponding to coherence requirements. Thus, in the meteorological problem with $q_1 = q(A_1) = 0.4$, $q_2 = q(A_2|A_1) = 0.8$ and $q_3 = q(A_1A_2) = 0.2$ we will have $\pi_3 = \pi_1\pi_2$. A possible set of reconciled values is given by $\hat{\pi}_i = E(\pi_i|q)$, the means of the distribution $p(\pi|q)$. In statistical language the $\hat{\pi}$'s are estimates of the π's. The precision of the reconciled value may be described by the inverse of the variance of π_i given q. Notice that, in general, the reconciled values will not be independent.

A special case arises when one or more of the events in A are of particular interest--we call them target events --and the other events are introduced in order to increase the precisions. It is then only necessary to calculate the marginal distribution of those π's which refer to the target events. We call this procedure extension of the conversation --from the target events to other events. The increase in precision essentially results from the increased exposure S has to the constraints of coherence when he contemplates many events. Example 2 above provides an illustration of this procedure. We believe that this method could be of considerable value in improving probability assessments. The calculations are described in detail in Section 4.3 below.

2.3 The External Approach

Here N is concerned only with q and A, so that S plays no role except as the provider of data q for N to update his probabilities of A. We have [2]

(4) $$p(q|A) = \Sigma_{\pi} p(q|\pi)p(\pi|A),$$

using the independence condition (1), and p(A) is available directly providing a complete probabilistic description of q and A. By Bayes' Theorem we have

(5) $$p(A|q) \propto p(q|A)p(A)$$

as N's appraisal of the uncertain events after S has reported his assessments q.

The meteorological situation provides an illustration in which the meteorologist, S, has announced a probability of rain tomorrow as q = 0.8; what is N's probability in the light of this information? In particular, should N use S's stated value? The external approach could be useful where N is a decision maker who requires a probability for A in order to take action. He may consult an expert, or a group of experts, each of whom reports his probability for A and the decision maker has to reach an overall judgment (see Example 2).

[2] We use Σ_{π} to denote a summation over π. In application below it will be an integration over real vector space.

Notice that although both approaches depend on the common core of distributions I-III above, $p(A)$, $p(\pi|A)$ and $p(q|\pi)$, they can be used without N determining all three. Thus, in the internal approach, $p(\pi)$ may be assessed directly, rather than through (2), when combination with $p(q|\pi)$ in (3) gives the required result. Similarly, in the external approach, $p(q|A)$ may be assessed directly, rather than through (4), and combined with $p(A)$ in (5) to produce the result. If these ideas are adopted, the internal approach is seen to be simpler than the external one because it avoids the second of the core distributions, $p(\pi|A)$, or the derived $p(q|A)$. Both these distributions are relatively unfamiliar in comparison with I and III, which are respectively a "prior" and a likelihood. However, II requires a judgment by N of what S believes about A (in the case of π) or will say he believes (for q) for each constituent event in A.

In the weather example, N has to consider what the meteorologist might say about rain tomorrow both on those occasions when it will rain, and on those when it will not. Presumably, for a reputable weather forecaster, the former will be higher than the latter. The distribution $p(\pi|A)$ and $p(\pi|\bar{A})$ are measures of the quality of the expert, S. Note that $p(q|A)$ is related to the idea of calibration discussed earlier. To see the relation, consider a sequence of occasions on which S asserted a value q for the probability of an event and let $f(A|q)$ denote the relative frequency with which the events in the sequence occur. Clearly, $f(A|q)$ is closely related to $p(A|q)$ which, in turn, is related to $p(q|A)$ by the one-dimensional form of (5). Recall that a person is calibrated if $f(A|q) = q$; that is, if a proportion q of the events to which he has assigned a probability q actually occur. Indeed, if $p(A|q) = q$, N will take S's announced value, q, to be his probability for A.

There is another relationship between the internal and the external approaches. Recall that their end-products are $p(\pi|q)$ and $p(A|q)$, respectively. The former provides N's judgment about S's coherent values, and the latter gives N's assessment of the external world in the light of S's information. In the internal approach, we proposed that $\hat{\pi} = E(\pi|q)$ might be used as reconciled values for S's statements. This raises the question of whether $\hat{\pi}(A) = p(A|q)$, that is, whether N could use the reconciled values for decision making. To answer this question note that

$$(6) \qquad p(A|q) = \Sigma \pi p(A|\pi)p(\pi|q),$$

using the independence condition (1). This is $\hat{\pi}(A)$ if $p(A|\pi) = \pi$. Hence the reconciled value will agree with N's probability if N would have been prepared to use S's coherent probability for A had he known it.

We now add a remark about the internal approach in its direct form using $p(\pi)$ and $p(q|\pi)$ to obtain $p(\pi|q)$. In this form, π plays the role of a set of parameters, q is data, and the "prior" $p(\pi)$ is updated by the likelihood $p(q|\pi)$ to give a posterior $p(\pi|q)$. The use of reconciled values $\hat{\pi} = E(\pi|q)$ with their associated precisions is closely related to the method of least squares. Suppose the data $q_1 \ldots q_m$ are, given the π's, independent and normally distributed about the π's with constant variance; suppose further that the prior for the π's is, relative to this likelihood, rather smooth. It then follows by standard theory that the π's given the q's are also normal. The means are the least-squares estimates obtained by minimizing $\Sigma_i (q_i - \pi_i)^2$ over the π's, subject to the coherence constraints. (This result is exact if the constraints are linear and will be approximate for non-linear constraints.) The matrix of second

derivatives of the sum of squares at the minimum when inverted gives approximate variances and covariances. If the q's are not independent but have a general normal distribution, then a weighted sum of squares and products replaces the direct sum above. Because the q's are bounded, they cannot be normally distributed so that it may be preferable to convert them to log-odds, $\ln[q/(1-q)]$. The whole argument then goes through with log-odds instead of probabilities, both for the q's and the π's.

This technique is particularly simple and is the most usable method we have for reconciliation of probabilities. It is discussed in some detail in Section 4.0. Notice that all it requires, in addition to S's provision of the data, q, are their variances and covariances since these completely describe the normal likelihood, and the prior is assumed to be "flat." In our formulation, the *likelihood* $p(q|\pi)$ has been thought of as N's but S may provide his own second moments. If these are used directly by N, our argument is unaffected.

3.0 TECHNICAL DEVELOPMENTS

In this section, we take the general model of the previous section, insert specific forms for the core distributions and calculate other distributions of interest. To avoid excessive technicalities, we confine ourselves to the very simplest cases and only aim to demonstrate the feasibility of the model. In Section 4.0, by specializing and using least-squares ideas, we come nearer to results of practical use. This section deals with two special cases in which all the calculations involved in the model can be displayed explicitly.

In the first case, there is a single event, A, for which S has true probability $\pi(A)$, or simply π, and for which he reports the single value $q(A)$, or q. Hence all quantities are one dimensional. With one value reported, there is no opportunity to use coherence; nevertheless, reconciliation, in the sense of calculating $\hat{\pi}$, might be appropriate depending on N's judgment about S expressed through the core distributions I-III. In any case, N may need to calculate his probability of A in the light of S's reported value q, as is the case when the weather forecaster reports the probability of rain tomorrow to be q. Another example arises in medical diagnosis: a physician, S, reports the probability q that N has appendicitis. What is N's probability for appendicitis?

To specify the model completely, we need to describe $p(A)$, $p(\pi|A)$, $p(\pi|\bar{A})$ and $p(q|\pi)$. It is convenient and sensible to work with log-odds, (see above) rather than with probabilities. To economize on notation, we use q and π to denote log-odds rather than probabilities. The log-odds for A is written $lo(A)$. Clearly, the general theory so

far discussed is unaffected by this change.

There are two reasons for changing to log-odds. First, it is necessary to handle bivariate distributions, and the normal is computationally the most attractive; so variables, like log-odds, with infinite ranges are to be preferred. Second, it is more reasonable to suppose that the measurement error has constant variance when expressed in log-odds rather than in probabilities, since values of the latter are, in absolute terms, more precisely assessed when near 0 or 1 than when around 1/2.

We therefore suppose that the last of our core distributions $p(q|\pi)$ is, in log-odds, normal with mean π and constant variance, σ^2, abbreviated to $N(\pi,\sigma^2)$. That is, N views S's measurement of log-odds as unbiased with constant variance. With $p(A) = \alpha$, say, we have only to specify $p(\pi|A)$ and $p(\pi|\bar{A})$. Suppose the former is $N(\mu_1,\tau^2)$. That is, if it really is going to rain tomorrow, then N expects S to have log-odds μ_1, with standard deviation τ. Similarly, suppose the latter (applying to the case of no rain) is $N(\mu_2,\tau^2)$. Presumably, for a good forecaster, $\mu_2 < 0 < \mu_1$. (Log-odds of zero correspond to a probability of 1/2.)

A special case arises when S is thought to be just as good when A is true as when it is false: then $\mu_2 = -\mu_1$, for when A is true, his probability for A would then be expected to be the same as that for $\bar{A}$ when $\bar{A}$ is true, but $p(\bar{A}) = 1-p(A)$ and hence $lo(\bar{A}) = -lo(A)$. This may be appropriate in the meteorological case but not in the appendicitis example, where it is easier to diagnose a real case of appendicitis than it is other sources of abdominal pain, so that perhaps $|\mu_2|<\mu_1$. The variance has been supposed the same for A and for $\bar{A}$: it would be possible to handle the more general case, but the algebra is untidy and little extra insight is gained. Notice how $p(\pi|A)$ and $p(\pi|\bar{A})$

together describe N's opinion of S as a probability appraiser. The values are related to the errors of the two kinds studied in statistics, or to the false positives and false negatives considered in medicine.

Notice that σ^2 and τ^2 describe two quite different aspects of variability in the situation. The former gives N's view of how far S's stated value q will differ from π, his coherent value, and corresponds to a measurement error. On the other hand, τ^2 describes N's view of S as a probability appraiser, specifying the variability in S's coherent value when A is true (or false).

In summary, the model is as follows:

$$\text{I, } p(A) = \alpha$$

$$\text{II, } p(\pi|A) \sim N(\mu_1, \tau^2)$$

$$p(\pi|\bar{A}) \sim N(\mu_2, \tau^2)$$

$$\text{III, } p(q|\pi) \sim N(\pi, \sigma^2).$$

Consider first the internal approach. We have, equation (2),

$$p(\pi) = p(\pi|A)p(A) + p(\pi|\bar{A})p(\bar{A})$$

$$= \alpha N(\mu_1, \tau^2) + (1-\alpha)N(\mu_2, \tau^2), \tag{6}$$

a weighted average of two normals. For sufficiently small values of τ^2 (if $\alpha = \frac{1}{2}$, $\mu_2 = -\mu_1$, the condition is $\tau^2 < \mu_1^2$), the distribution is bimodal. Thus, good meteorologists (with small standard deviations) typically quote probabilities for rain tomorrow which are either high (when it is going

to rain) or low (when not), and only rarely are they so undecided as to give values around 1/2, log-odds of zero.

It is convenient to write $p(\pi|A) = p_1(\pi)$ and $p(\pi|\bar{A}) = p_2(\pi)$. Then, by Bayes Theorem,

$$\begin{aligned} p(\pi|q) &= p(q|\pi)p(\pi)/\Sigma_\pi p(q|\pi)p(\pi) \\ &= \frac{p(q|\pi)\{\alpha p_1(\pi) + (1-\alpha)p_2(\pi)\}}{\alpha p_1(q) + (1-\alpha)p_2(q)} \\ &= p_1(\pi|q)\alpha' + p_2(\pi|q)\ (1-\alpha') \end{aligned}$$

where $p_i(q) = \Sigma_\pi p(q|\pi)p_i(\pi)$, $p_i(\pi|q) = p(q|\pi)p_i(\pi)/p_i(q)$ and $\alpha' = \alpha p_1(q)/\{\alpha p_1(q) + (1-\alpha)p_2(q)\}$.

From standard normal theory, $p_i(\pi|q)$ is normal with mean $wq + (1-w)\mu_i$, $w = \tau^2/(\sigma^2+\tau^2)$ and $p_i(q) \sim N(\mu_i,\sigma^2+\tau^2)$, so that $p(\pi|q)$ is also a mixture of normals. The mean is

$$\text{(8)}\quad \hat{\pi} = E(\pi|q) = \alpha'[wq + (1-w)\mu_1] + (1-\alpha')\ [wq + (1-w)\mu_2]$$
$$= wq + (1-w)\ [\alpha'\mu_1 + (1-\alpha')\mu_2].$$

This is the reconciled value of π on the basis of the stated value q. It is near q if w is near one, that is, if τ^2 is much greater than σ^2. A large value of τ means that $p(\pi)$ [equation (6)] has large spread and that the likelihood $p(q|\pi)$ with smaller spread σ is dominant. This is the situation discussed in connection with least-squares methods in Section 2.3. Consequently, if S has small measurement error compared with appraisal error, the reconciled value will be essentially the stated value, and no change will occur. The variance, and hence the precision, of the reconciled value, $var(\pi|q)$ can be found. It is an untidy

expression; but in the case where τ^2 is much larger than σ^2, it is approximately equal to σ^2.

A numerical example when τ^2 is of the same order as σ^2, so that reconciliation away from q may take place, is instructive. Suppose $\alpha = 1/2$, $\mu_1 = -\mu_2 = 1.0$ (so that S is equally competent whether A is true or false), $\tau = 1.0$ and $\sigma = 0.5$. At two standard deviations, when A is true, S is anticipated to have true log-odds between -1.0 and 3.0 (probabilities between .27 and .95) and when false between -3.0 and 1.0 (.05 and .73 for probabilities). But the reported odds can differ by as much as 1.0(=2σ) from the true values. (A probability of 0.5 can be reported anywhere in the range .27 to .73). Then w = 0.8 so that in $\hat{\pi}$, equation (8), 80% of the weight goes on q and 20% depends on the anticipated performance of S. If q = 0.7 (stated probability of 0.67), the weight $\alpha' = 0.75$ and $\hat{\pi}$ is 0.66 (a reconciled probability of 0.66). If q = 1.5 (probability 0.82), the weight $\alpha' = 0.92$ and $\hat{\pi}$ is 1.37 (probability 0.80). Both probabilities are lowered slightly, the mean of the prior distribution of π being zero. But the changes are small, and even here, where τ is only twice σ, the approximation that assumes τ is large is not unreasonable. In this case, reconciliation is merely allowing for the effect of measurement error.

As an intermediary between the internal and external approaches, we can consider $p(A|\pi)$. We saw just after equation (6) that N could use the reconciled value $\hat{\pi}$ as his probability for A if $p(A|\pi) = (1 + e^{-\pi})^{-1}$, remembering that π is now a log-odds. We have

$$p(A|\pi) = p(\pi|A)p(A)/p(\pi)$$

$$= \frac{\alpha p_1(\pi)}{\alpha p_1(\pi) + (1-\alpha)p_2(\pi)}$$

from (6). Simple algebra shows the log-odds to be

$$(9) \qquad lo(A|\pi) = lo(A) - (\mu_1^2 - \mu_2^2)/2\tau^2 + (\mu_1 - \mu_2)\pi/\tau^2 .$$

For this to be π, two conditions must be fulfilled, namely $(\mu_1-\mu_2)/\tau^2 = 1$ and $\ln\frac{\alpha}{1-\alpha} = (\mu_1^2-\mu_2^2)/2\tau^2$. These are equivalently

$$(10) \qquad (\mu_1-\mu_2)/\tau^2 = 1 \text{ and } 1/2(\mu_1+\mu_2) = \ln\frac{\alpha}{1-\alpha} .$$

The second condition says that the prior log-odds have to equal the average performance. This is satisfied in the special case $\mu_2 = -\mu_1$ when $\alpha = ½$. The first condition says that the difference between the expected performances under the two conditions, A and $\bar{A}$, must equal the appraisal variance τ^2. Thus, a complete matching of N and coherent-S when considering A depends on a rather complex combination of prior views by N and S's appraisals. A change in the former would affect this agreement. If N judges A to be as likely as not, $\alpha = ½$, the conditions for agreement are that $\mu_2 = -\mu_1$, so that S performs equally well under A as $\bar{A}$, and $2\mu_1/\tau^2 = \lambda$, say is one. No simple interpretation of this final requirement is known to us.

Consider next the external approach. It is immediate from the core distributions II and III that

$$(11) \qquad p(q|A) \sim N(\mu_1, \sigma^2+\tau^2) \text{ and } p(q|\bar{A}) \sim N(\mu_2, \sigma^2+\tau^2),$$

so that the replacement of π (in II) by q here merely replaces τ^2 by $\sigma^2+\tau^2$. Consequently, $p(A|q)$ is given by the same result as (9) but with τ^2 everywhere replaced by $\sigma^2+\tau^2$. The "least-squares" case where τ^2 is large compared with σ^2 makes the two quantities $p(A|\pi)$ and $p(A|q)$ nearly equal.

In the numerical illustration with $\alpha = 1/2$, $\mu_1 = \mu_2 = 1.0$, $\tau = 1.0$ and $\sigma = 0.5$, so that $\lambda = 2$, the log-odds given q are, from (9) with $\sigma^2+\tau^2$ replacing τ^2 and q for π, equal to $(\mu_1-\mu_2)q/(\sigma^2+\tau^2)$, or 1.6 times q. Hence q of 0.7 (probability 0.67) is raised to q of 1.12 (probability 0.75) and a q of 1.5 (probability 0.82) to one of 2.4 (probability 0.92). These changes are much larger than those caused by reconciliation. The intuitive explanation for this is most easily seen by considering the higher value of q, 1.5. Such a value is much more likely to have arisen from the distribution when A is true, $p_1(q) \sim N(1,1.25)$ rather than from the distribution when A is false, $p_2(q) \sim N(-1, 1.25)$. This is described by α' which is here 0.92, so that there is a 92% chance that q came from A. Consequently, N can substantially increase S's stated value. Notice how this change depends heavily on S as a probability appraiser whereas reconciliation does not, $p(\pi|A)$ and $p(\pi|\bar{A})$ enter only through $p(\pi)$. If τ increases, the effect on N will become less. In an extreme case where τ and σ tend to zero, N will shift S's stated value to either 1 or 0. A meter-ologist who always says 0.6 chance of rain on rainy days and 0.4 when it is to be dry entitles N to be sure of rain when 0.6 is announced.

The second example differs from the first only in that S assesses both the probability of A and of $\bar{A}$, supplying $q(A)$ and $q(\bar{A})$. These we denote by q_1 and q_2 respectively. There is now a possiblity for incoherence since they may not add to 1. With rather general normal distributions, the algebra becomes complicated; but it is possible to demon-strate the main points if we specialize somewhat. Specifi-cally, we suppose (the arguments still being log-odds)

I: $p(A) = \frac{1}{2}$,

II: $p(\pi|A) \sim N(\mu,\tau^2)$

$p(\pi|\bar{A}) \sim N(-\mu,\tau^2)$

(these are as before with $\alpha = \frac{1}{2}$ and $\mu_2 = -\mu_1$).

For the last stage, we need to describe the joint distribution of q_1 and q_2. This we suppose:

III: $p(q_1|\pi) \sim N(\pi,\sigma^2)$ and

$p(q_2|\pi) \sim N(-\pi,\sigma^2)$,

these being independent, given π.

The last condition says that both q_1 and $q_2 = q(\bar{A})$ are unbiased, of equal precision and are uncorrelated. This imples that, for all π, $E(q_1+q_2|\pi) = 0$, so that N expects S to be coherent. A generalization would allow for bias. Another generalization would permit q_1 and q_2 to be correlated. For example, S may always have $q_1+q_2 = 0$, in which case they are perfectly negatively correlated and we are back to the first example.

In the internal approach, the calculation of $p(\pi)$ is as before, equation (6), with its possible bimodal form. The revised value of this distribution $p(\pi|q_1,q_2)$ is of the same character as before, equation (7); a mixture of two normal distributions separately derived from $p_1(\pi) = p(\pi|A)$ and $p_2(\pi) = p(\pi|\bar{A})$. Standard normal theory calculations show that $p_1(\pi|q_1,q_2)$ has mean

$$\left\{\frac{(q_1-q_2)}{\sigma^2} + \frac{\mu}{\tau^2}\right\} \Big/ \left\{\frac{2}{\sigma^2} + \frac{1}{\tau^2}\right\}$$

with variance $\{2/\sigma^2 + 1/\tau^2\}^{-1}$: and that $p_2(\pi|q_1,q_2)$ is of the same form with $-\mu$ for μ. The weights are $w' = p_1(q_1,q_2)/\{p_1(q_1,q_2) + p_2(q_1,q_2)\}$ and $1-w'$. Thus the reconciled value is

$$\text{(12)} \qquad \hat{\pi} = E(\pi|q_1,q_2) = \frac{(q_1 - q_2)/\sigma^2}{2/\sigma^2 + 1/\tau^2} + \frac{\mu(2w'-1)/\tau^2}{2/\sigma^2 + 1/\tau^2} .$$

For the case where τ^2 is large compared with σ^2, this is about $\frac{1}{2}(q_1 - q_2)$. This simple form is interesting: it is $\frac{1}{2}\{q_1 + (-q_2)\}$, the average of q_1 and $(-q_2)$ which are effectively two statements made by S about A, and not one about A and one about $\bar{A}$. The first is direct; the second was originally q_2 for $\bar{A}$, so $-q_2$ for A if coherent.

A numerical example may prove illuminating. Suppose S gave probabilities of 0.6 for A and 0.3 for $\bar{A}$. The corresponding log-odds are $q_1 = 0.41$ and $q_2 = -0.85$. Then $1/2\ (q_1-q_2) = 0.63$, giving a reconciled value for S's coherent probability for A of 0.65, and of course 0.35 for $\bar{A}$.

The variance of the reconciled value of the log-odds is $\frac{1}{2}\sigma^2$, for large τ^2; one half the variance, twice the precision, of the separate assessments. Suppose, in illustration, $\sigma = 0.2$, so that at two standard deviations the log-odds for A could lie in the range $0.41 + 0.4$, or in probability terms (0.50,0.69). Similarly, the range of probabilities for $\bar{A}$ is (0.22, 0.39). The consequent range

for the reconciled probability of A, $\hat{\pi}$ is (0.59, 0.71).

The calculation of $p(A|\pi)$ is as before with no alteration, and (9) simplifies to give $lo(A|\pi) = 2\mu\pi/\tau^2 = \lambda\pi$ in the earlier notation.

For the external approach, we have first to combine II and III to obtain $p(q_1,q_2|A)$ and $p(q_1,q_2|\bar{A})$. The former is clearly a bivariate normal distribution with means $(\mu,-\mu)$, variances $\sigma^2 + \tau^2$ and covariance $-\tau^2$. Conditional on $\bar{A}$, the distribution is the same with $-\mu$ for μ. Direct calculation using Bayes Theorem shows that

$$(13) \qquad lo(A|q_1,q_2) = 2\mu(q_1-q_2)/(\sigma^2+2\tau^2)$$

If τ^2 is large in comparison with σ^2, this is about $2\mu\hat{\pi}/\tau^2$. If $\tau = 1.0$ (taken as large in comparison with $\sigma = 0.2$) and $\mu = 1.0$, our numerical illustration gives a value for the log-odds of 1.20, corresponding to a probability of 0.77. This is a substantial increase from a stated probability of 0.6, reconciled to 0.65 to allow for incoherence. The reason is, as in the first example, that N judges that the stated values, 0.6 and 0.3, have more likely arisen from A, than $\bar{A}$, so that he increases the value for $p(A|q_1,q_2)$.

These two examples demonstrate how the ideas work out in simple cases and show that the internal approach is simpler than the external one, particularly when least-squares procedures are used. The difficulty with the external approach lies in the use of the measures of S's ability as a probability appraiser, $p(\pi|A)$, which do not enter into the other apparoach. Nevertheless, there are problems where the external approach is essential, as when N has separate

opinions from several experts and needs to use them to assess his own probability for A. We pass in the next section to a consideration of the internal approach using least squares.

4.0 LEAST-SQUARES PROCEDURES WITH THE INTERNAL APPROACH

We begin with a general description of the procedure. We are concerned only with q and π, S's stated and coherent values for some events. N's opinion about π is diffuse so that $p(\pi)$ is approximately constant. In the simpler situations, each element q_i of q is approximately normally distributed about π_i, the corresponding element of π, with constant variance σ^2 say, these being independent. Under these circumstances, the reconciled values $\hat{\pi}_i = E(\pi_i|q)$ are given approximately by the values of the π's that minimize

$$\Sigma_i (q_i - \pi_i)^2$$

subject to any constraints on the π's that coherence imposes. In other situations, the q's will be correlated and have different variances, when a quadratic form

$$\Sigma w_{ij} (q_i - \pi_i)(q_j - \pi_j)$$

will be minimized subject to the constraints, the w's being weights, the elements in the inverse of the dispersion matrix of the q's. Minimization is performed by equating the first derivatives to zero. The matrix of second derivatives at the minimum, when inverted, provides the approximate variances, $\text{var}(\pi_i|q)$, and covariances of the reconciled values. Throughout this argument, the π's and q's may be probabilities or some suitable transform of them, such as log-odds.

The whole procedure is straightforward except for one difficulty: the constraints imposed by coherence are typically non-linear, and the resulting equations are therefore non-linear. There is no simple resolution of this

difficulty. The power of the probability calculus lies in the ability of probabilities to combine both additively and multiplicatively. If linearity of the latter is imposed by taking logarithms, the linearity of the former is destroyed. It is, therefore, a fundamental difficulty but one that can sometimes be alleviated by suitable approximations.

The coherent probabilities $\pi_o (1 \le i \le n)$ will themselves be functions of $k \le n$ probabilities that can each coherently receive any value in the unit interval irrespective of the other. For example, with two events, A_1, A_2, where S has provided $q_1 = q(A_1)$, $q_2 = q(A_2|A_1)$ and $q_3 = q(A_1A_2)$, then if $\theta_1 = \pi(A_1)$ and $\theta_2 = \pi(A_2|A_1)$, we have $\pi_1 = \theta_1$, $\pi_2 = \theta_2$ and $\pi_3 = \theta_1\theta_2$ where $0 < \theta_1, \theta_2 < 1$, $k = 2$, $n = 3$. If $k < n$, then coherence imposes constraints on the π's. Denote, as in the example, these basic probabilities by $\theta_1, \theta_2, \ldots \theta_k$. Then confining attention to the uncorrelated, equal weights case (generalizations being clear), the expression to be minimized with respect to the θ's is

$$(14) \qquad \Sigma[q_i - \pi_i(\theta_1, \theta_2, \ldots \theta_k)]^2.$$

Differentiation with respect to θ_j gives

$$(15) \qquad \Sigma_i[q_i - \pi_i(\hat{\theta})] \; \partial\pi_i/\partial_j = 0 \quad .$$

The usual way to solve these equations in the non-linear case is to suppose we have some first guess θ_o at $\hat{\theta}$--for example, the values suggested by $q_1, q_2, \ldots q_k$--and expand $\pi_i(\hat{\theta})$ in a Taylor series about θ_o retaining only the first term. The resulting equations are linear in the discrepancy $\hat{\theta}-\theta_o$ and may therefore be solved.

Numerically these procedures may be iterated to obtain improved solutions. In what follows, we shall concentrate

on analytical solutions, either exact or approximate, which help us to understand the methods, recognizing that for practical implementation in complex situations, computer software will have to be provided. This provision should not be difficult since many minimization procedures for non-linear functions are available.

We now pass to the consideration of several special cases. In any application of the least-squares ideas, we have to specify whether the calculation is in terms of probabilities directly, or in some function thereof, such as log-odds. We refer to this as the choice of metric: probability metric, log-odds metric, etc. The reason for preferring one metric to another is, as before, that the variance of the q's may be judged to be more reasonably constant in one metric than in another.

4.1 Partition: General Metric

Here $(A_1, A_2, \ldots A_n)$ is a partition, and S provides probabilities $q_i = q(A_i)$, $1 \le i \le n$ for $\pi_i = \pi(A_i)$ with the single[1] coherence constraint $\Sigma\pi_i = 1$. For a general metric $F(\cdot)$ -- for log-odds, $F(t) = \ln[t/(1-t)]$ - the expression to be minimized is $\Sigma[F(q_i) - F(\pi_i)]^2$ subject to the constraint $\Sigma\pi_i = 1$. The constraint can either be incorporated as above, with $\theta_i = \pi_i$ for $i<n$ and $\pi_n = 1 - \Sigma_{i=1}^{n-1}\theta_i$, or by using a Lagrangian. For the probability metric, the equations are linear and have the exact solution

$$(16) \qquad \hat{\pi}_i - q_i = n^{-1}(1 - \Sigma q_j)$$

with

$$(17) \qquad \mathrm{var}(\hat{\pi}_i) = (1 - n^{-1})\sigma^2,$$

---

[1] It is assumed throughout that S always gives values for the q's that lie in the unit interval.

σ^2 being the variance of each value stated by S. The improvement in precision due to coherence is only appreciable when n is small; for n = 2 it doubles, as we saw in the previous section. The form of (16) is worth a comment since it generalizes. The adjustment, $\hat{\pi}_i - q_i$, of the stated value to a reconciled value is a multiple, here n^{-1}, of the degree of incoherence, that is, how far Σq_j differs from its coherent value, 1. Here each q_i is altered by the same amount.

A subject was asked to assess the probabilities that his ultimate death would fall under one of the five categories: cancer, heart disease, stroke, other natural, unnatural (accident, suicide, etc.). He gave the values given in the second column of Table 1.[2] The total is 0.83, exhibiting a fair degree of incoherence perhaps caused by the unpleasant nature of the events. The reconciled values are each increased by 0.17/5 = 0.034 and are given in column six.

In other metrics the minimization equations

$$[F(q_i) - F(\hat{\pi}_i)]F'(\hat{\pi}_i) = \lambda$$

in Lagrange form are non-linear. Using the device suggested above, a first guess for π_i might be q_i with the result that

$$\hat{\pi}_i = q_i - \lambda/F'(q_i)^2. \tag{18}$$

[2]The last column gives the frequencies (unknown to S) for U.S. residents. These values are not to be interpreted as the correct probabilities. For example, a person with a known history of heart trouble would have a higher probability than the frequency value--a point we return to later.

Cause of Death	Stated Values q_i	Ranges for q_i	Weights	Reconciled Values $\hat{\pi}_i$	$\hat{\pi}_i$ Unweighted	$\hat{\pi}_i$ Log-odds	U.S. Frequencies
Cancer	.25	.15-.35	25	.28	.28	.31	.18
Heart Disease	.20	.05-.40	8	.25	.23	.24	.40
Stroke	.10	.05-.25	25	.12	.13	.11	.11
Other Natural	.20	.05-.45	6	.27	.23	.24	.25
Unnatural	.08	.01-.15	51	.09	.11	.09	.06
	.83						

Table 4-1

Subject's Probabilities for Various Causes of Death

With the log-odds metric $F'(x) = x^{-1}(1-x)^{-1}$ so that

$$(19) \qquad \hat{\pi}_i = q_i - \lambda q_i^{\,2}(1-q_i)^2$$

providing a larger correction when q_i is near 1/2 than elsewhere. In the medical example, the changes upwards are 0.06, 0.04, 0.01, 0.04 and 0.01, instead of a constant 0.034 in the probability metric. The reconciled values are given in the penultimate column of Table 4-1.

The reason for using other than the probability metric is the fact that not all stated probabilities may have the same errors associated with them. An alternative to changing the metric is for either S or N to provide the precision associated with each value. The precision of an assessment is closely related to its stability, i.e., the degree to which it varies upon further reflection. The assessment of the probability that a coin will come up heads, for example, is very precise: it is not likely to depart from 0.5. The assessment of the probability that a thumb-tack will fall on its head, however, is much less precise. Different ways of thinking about the problem could yield widely different assessments.

To elicit the precision associated with each of the q's, we asked the subject to quote a range of values for each assessed probability (see column 3 of Table 4-1). The quoted ranges were interpreted as multiples of standard deviations and weights used inversely proportional to the variances (column 4). We now have to minimize $\Sigma w_i(q_i-\pi_i)^2$ with the result that

$$(20) \qquad \hat{\pi}_i - q_i = (1 - \Sigma q_j)/w_i \Sigma w_j^{-1} \ .$$

The results for the medical example are given in the fifth column of Table 4-1.

4.2 Sub-partition: Probability Metric

A problem closely related to that described in the previous subsection arises where S assesses the probabilities for a number of exclusive, but not exhaustive, events $A_1, A_2, \ldots A_n$, giving values $q_1, q_2, \ldots q_n$, and also assesses the probability of their union, citing q'. Incoherence arises if $q' \neq \Sigma q_i$. An example arises in the same medical context where S gave values for four natural causes of death (cancer, heart disease, stroke and other) and also for the probability of his dying a natural death. With $\pi_i = \pi(A_i)$ and weights w_i, the problem is to minimize,

$$(21) \qquad \Sigma w_i (q_i - \pi_i)^2 + w(q' - \Sigma \pi_i)^2 .$$

The resulting equations are linear with solutions

$$(22) \qquad \pi_i - q_i = (q' - \Sigma q_j)/w_i \{\Sigma w_j^{-1} + w^{-1}\} .$$

Notice again the occurrence of the measure of incoherence $q' - \Sigma q_j$. If, in the medical example, the four probabilities for the types of natural death had been as before, and S had quoted 0.92 as the probability for natural death (so that he was coherent with his original 0.08 for unnatural death), then the revision is exactly as before. An alternative way of dealing with sub-partitions using conditional probabilities is described below.

Notice that in both these subsections it has been assumed that the correlations between S's different values are all zero. The most likely form of correlation is negative, the subject raising one probability and consequently lowering another. Indeed, a coherent S will have correlation $-n^{-1}$, giving $q' = \Sigma q_i$.

4.3 External Conditioning: Probability Metric

In Section 2.2 we mentioned the possibility of having some target events whose probabilities are required and that in order to assess these values, the conversation is extended to other events. Here we consider the simplest case of a single target event A and the extension to another event, denoted X to distinguish it clearly from a target event. Suppose S is sufficiently conscious of coherence so that his assessments for any pair of complementary events add to unity. Suppose further that S assesses $\pi(A)$, $\pi(A|X)$, $\pi(A|\bar{X})$, and $\pi(X)$. Here, the opportunity for incoherence arises since $\pi(A) = \pi(A|X)\pi(X) + \pi(A|\bar{X})\pi(\bar{X})$ may not hold for the corresponding q's. Our hope is that the reconciled value $\hat{\pi}(A)$ will be an improvement over the raw assessment $q(A)$, and we therefore wish to evaluate the precision of the reconciled value. The coherence constraint above is non-linear; to avoid this let us *suppose* $\pi(X)$ *is known*[3] so that $q(X) = \pi(X)$. We later generalize to the case where this is also in error. The process will be called <u>external conditioning</u>, since the probability of the target event is considered conditional on some external event, X.

Since no new difficulties arise by generalizing, we consider a partition into $X_1, X_2, \ldots X_n$ and not just X and $\bar{X}$, $n = 2$. Write $q(A|S_i) = q_i$, $q(A) = q'$, the π's analogously and $q(X_i) = \kappa_i = \pi(X_i)$, with $\Sigma\kappa_i = 1$. The coherence constraint is that $\pi(A) = \Sigma\pi(A|X_i)\pi(X_i)$ or $\pi' = \Sigma\pi_i\kappa_i$, and we have to minimize

---

[3]An example is where X is the event that a team wins the toss at the beginning of a contest. It would then generally be agreed that $q(X) = \pi(X) = p(X) = 1/2$. The event A could then be that the team wins the contest.

(23) $\Sigma(q_i-\pi_i)^2 + (q'-\Sigma\pi_i\kappa_i)^2,$

on assuming equal weights and zero correlations for the q's.

The result, for the target event A, is that

(24) $$\hat{\pi}' - q' = \frac{\Sigma q_i\kappa_i - q'}{1 + \Sigma\kappa_i^2} .$$

In words, the correction to q' for incoherence is the departure from coherence, $\Sigma q_i\kappa_i-q'$, divided by $(1+\Sigma\kappa_i^2)$. This divisor is the variance of the departure, assuming all q's have unit variance. The precision of the reconciled value is therefore easily found to be

(25) $$1 + (\Sigma\kappa_i^2)^{-1}$$

times the precision of each stated value.

There is another simple way of looking at $\hat{\pi}$ that generalizes. In effect, S has provided two values for $\pi(A)$; q' directly and $\Sigma q_i\kappa_i$ indirectly. These have variances σ^2 and $\Sigma\kappa_i^2\sigma^2$, where σ^2 is the variance of each statement by S. Taking a weighted average of these two values with weights inversely proportional to the variances, we obtain $\hat{\pi}$ as given by (24).

Returning to a single event X, the case n = 2, the precision of $\hat{\pi}[A|q(A),q(A|X),q(A|\bar{X})]$ is $1 + [\pi(X)^2 + \pi(\bar{X})^2]^{-1}$ relative to that for q(A). The improvement is remarkable. If $\pi(X) = 1/2$, as in the example of a toss, the reconciled value has three times the precision of the original value. The extension of the conversation to include X has tripled the precision, and it would appear that the extension is important in assess probabilities. As we shall see below, the increase in precision is

considerably smaller when the assessments are positively correlated, as is likely to be the case in most applications.

It is easy to see in the general case that the precision (25) is maximized when all the κ's are equal, that is, $\kappa_i = n^{-1}$: it is best to choose a partition in which the constituent events are all equally likely. The precision for the reconciled value is then (n+1) times the precision for the original value. (This was the case with the example of winning the toss.) Expressed differently, the increase in precision due to refining a partition from n to (n+1) elements is equivalent to that that could be had from an independent assessment for the probability of the target event. Since such independence judgments are not available in practice, the principle of extension of the conversation can be used to play a similar role.

The result that among partitions of size n the precision is maximized when all probabilities in the partition are equal is a very special solution to what we call the design problem, that is, the problem of designing questions to be answered probabilistically by S in such a way that the expected precision of the reconciled value for the target event is maximized. It is analogous to the design problem in weighing objects, where the purpose is to choose the weighings in such a way as to maximize the precision of the final determination of weight. We do not discuss this problem in the present paper except, as here, when solutions arise as a by-product of our investigation of the principle of reconciliation. One cannot choose the best design until the reconciliation problem for that design has been solved.

4.4 External Conditioning: Log-Odds Metric

With this metric the constraint imposed by coherence is non-linear and resort has to be made to devices exemplified by (15). The use of the log-odds metric is equivalent to saying that the variance of each q is proportional to $\pi^2(1-\pi)^2$, or $q^2(1-q)^2$ approximately. We may therefore replace (23) by the weighted sum

$$\Sigma w_i (q_i - \pi_i)^2 + w(q' - \Sigma \pi_i \kappa_i)^2, \tag{26}$$

in generalization of (21), with $w_i \propto [q_i(1-q_i)]^{-2}$ and $w \propto [q'(1-q')]^{-2}$. The result of the minimization is that

$$\hat{\pi}' - q' = \left(\frac{\Sigma q_i \kappa_i - q'}{\frac{1}{w} + \Sigma \frac{\kappa_i^2}{w_i}} \right) \frac{1}{w} \tag{27}$$

And the variance is proportional to

$$\frac{1}{w} \sum \frac{\kappa_i^2}{w_i} \left(\frac{1}{w} + \sum \frac{\kappa_i^2}{w_i} \right)^{-1} . \tag{28}$$

To illustrate the reconciliation procedure for external conditioning, consider the assessment of the probability of an energy crisis in the United States during the next decade, denoted A. Let X denote the development of new, effective methods for the use of solar energy during the next decade. Suppose $q' = q(A) = 0.7$, $q_1 = q(A|X) = 0.5$, $q_2 = q(A|\bar{X}) = 0.8$, and $\kappa = q(X) = \pi(X) = 0.5$. With the probability metric, the reconciled value is 0.67 with precision three times that of each of the q's, as we saw in Section 4.3. With the log-odds metric, the weights are proportional to $w_1 = .0625$, $w_2 = .0256$, and $w = .0441$. The reconciled value is still 0.67, and the precision is slightly more than three times

that of the original value, 0.7. This suggests that the procedure is fairly robust to changes in the metric.

4.5 External Conditioning: Correlations

When S provides $q(A)$ and $q(A|X)$, he will often find himself considering the same factors in both assessments. Consequently, it may be reasonable to think that the errors he makes in these two assessments are correlated. We investigate the effect of this on external conditioning in the special case of a partition into X and $\bar{X}$, where $q(A|X)$ and $q(A|\bar{X})$ are equally correlated with $q(A)$. The weights attached to the two conditional assessments are also supposed equal. In other words, the situation is symmetric in X and $\bar{X}$. Only the probability metric is considered.

The notation for the variances and covariances involved is defined by giving the dispersion matrix of the assessed quantities, other than $\pi(X) = \kappa$ which we are still assuming has no error. This is

$$\begin{array}{l} q' = q(A) \\ q_1 = q(A|X) \\ q_2 = q(A|\bar{X}) \end{array} \qquad \begin{pmatrix} \sigma^2 & \rho\sigma\tau & \rho\sigma\tau \\ \rho\sigma\tau & \tau^2 & \delta\tau^2 \\ \rho\sigma\tau & \delta\tau^2 & \tau^2 \end{pmatrix}$$

with inverse

$$\begin{pmatrix} \tau^2(1-\delta^2) & -\rho\sigma\tau(1-\delta) & -\rho\sigma\tau(1-\delta) \\ -\rho\sigma\tau(1-\delta) & \sigma^2(1-\rho^2) & \sigma^2(\rho^2-\delta) \\ -\rho\sigma\tau(1-\delta) & \sigma^2(\rho^2-\delta) & \sigma^2(1-\rho^2) \end{pmatrix}$$

divided by $\sigma^2\tau^2(1-\delta)(1+\delta-2\rho^2)$. It is this matrix that provides the weights to be used in the quadratic form that has to be minimized. Thus it contains a term

$$-\rho\sigma\tau(1-\delta)(q'-\pi')(q_1-\pi_1)/\sigma^2\tau^2(1-\delta)(1+\delta-2\rho^2)$$

corresponding to the element in the first row and second column of the matrix. Here $\pi = \kappa\pi_1 + (1-\kappa)\pi_2$ by coherence. Tedious calculation shows that

$$(29) \qquad \hat{\pi}' - q' = \frac{[\kappa q_1 + (1-\kappa)q_2 - q']\sigma(\sigma-\rho\tau)}{K\tau^2(1-\delta) + \delta\tau^2 + \sigma^2-2\rho\sigma\tau}$$

where $K = \kappa^2 + (1-\kappa)^2$. With correlations ρ and δ equal to zero, this reduces to (27) with $n = 2$, $w_1 = w_2 = \tau^{-2}$ and $w = \sigma^{-2}$. It may also be obtained by taking a weighted average of $\kappa q_1 + (1-\kappa)q_2$ and q' with weights that depend on the variances and correlations. The variance of the reconciled value is given by

$$\mathrm{var}(\hat{\pi}) = \frac{\sigma^2\tau^2[K(1-\delta) + (\delta-\rho^2)]}{K\tau^2(1-\delta) + \delta\tau^2 + \sigma^2-2\rho\sigma\tau}$$

Let us return to the example of the energy crisis where $q' = 0.7$, $q_1 = 0.5$, and $q_2 = 0.8$, with $\kappa = 0.5$, and $\sigma^2 = \tau^2 = 1$. If $\rho = \delta = 0.5$, the reconciled value is 0.67 exactly as before, but the variance is decreased from 1.0, say, to 0.67; that is, the precision is only increased by 50% rather than tripled as before. This result reflects a general tendency for correlations to have little effect on the probabilities but a substantial effect on the precisions.

A comment on the assessment of correlations is in order. Recall that in order to elicit variances, we asked the subject to set upper and lower limits for each of his q's, and treated the width of these intervals as proportional to the respective standard deviations. A similar procedure can perhaps be used to elicit the correlation between, say, q_1 and q_2: the subject can be asked to set limits for q_1--assuming q_2 is fixed. The ratio of the width

of the interval for q_1 when q_2 is fixed to the width of the same interval when q_2 is allowed to vary, is a simple transform of the correlation coefficient. However, this task may be too demanding, and it appears difficult, although not impossible, to elicit from subjects meaningful judgments of correlations between assessments. Fortunately, the reconciled probabilities are fairly insensitive to the correlations, which affect primarily the precisions. Thus, instead of asking the subject to assess the correlations between his judgments, we could ask him only for upper and lower limits for these correlations. These values could then be used to obtain upper and lower bounds for the precisions of the reconciled probabilities.

4.6 External Conditioning: $\pi(X)$ Assessed

To help in our understanding of the situation, external conditioning has been considered only when the probability of the conditioning event is known, since this makes the analysis linear. We now pass to the general case supposing that all four probabilities are assessed with the same errors and with no correlations. As before, we write $q(A) = q'$. $q(A|X) = q_1$, $q(A|\bar{X}) = q_2$ and introduce $q(X) = q_3$. The π's correspond and the coherence constraint is that $\pi' = \pi_1\pi_3 + \pi_2(1-\pi_3)$. The function to be minimized is

$$(30) \qquad (q_1-\pi_1)^2 + (q_2-\pi_2)^2 + (q_3-\pi_3)^2 + (q'-\pi')^2.$$

It is possible to proceed using Taylor series expansion, but in this case a simpler approximation is available using the alternative way of interpreting $\hat{\pi}'$ in (24) as a weighted average. Here the two values for $\pi(A)$ cited by S are q' directly and $q_1q_3 + q_2(1-q_3)$ indirectly. The variance of the former is σ^2, say. The latter is non-linear, but its differential is $q_3\delta q_1 + (1-q_3)\delta q_2 + (q_1-q_2)\delta q_3$ so

that its variance is approximately $q_3^2 + (1-q_3)^2 + (q_1-q_2)^2$, times σ^2. The evaluations are independent. Taking the weighted average and rearranging to obtain the same form as before, we have

$$\hat{\pi}' - q' = \frac{q_1 q_3 + q_2(1-q_3) - q'}{1 + q_3^2 + (1-q_3)^2 + (q_1-q_2)^2} \tag{31}$$

The precision is

$$1 + [q_3^2 + (1-q_3)^2 + (q_1-q_2)^2]^{-1} \quad . \tag{32}$$

Both these results are approximate but probably adequate for most applications. The result for the precision is especially interesting since it shows, in comparison with the result for known $\pi(X)$, the reduction in precision due to uncertainty about $\pi(X)$. A straight use of the precision result for known $\pi(X)$, equation (25), would give, in the present notation $1 + \{q_3^2 + (1-q_3)^2\}^{-1}$, differing from (32) in not having the term $(q_1-q_2)^2$ in brackets. It is this term that produces reduction in precision. Hence, lack of knowledge of $\pi(X)$ is most critical when $q_1 \neq q_2$; that is, $q(A|X) \neq q(A|\bar{X})$. If $q_1 = q_2$ then it does not matter. This is intuitively sensible.

In the energy problem, we have as before: $q' = 0.7$, $q_1 = 0.5$, $q_2 = 0.8$, and $q_3 = 0.5$, except that q_3 is now uncertain. The reconciled value of π is 0.67 as before, illustrating the robustness of the reconciliation procedure. The precision is increased by the external conditioning from 1 to 2.70, instead of 3.0 as before. By using the full equations (30), it is possible to calculate the other reconciled values, which are $\hat{\pi}_1 = 0.52$, $\hat{\pi}_2 = 0.82$ and $\hat{\pi}_3 = 0.49$.

The extension to a general partition is straightforward.

4.7 External Conditioning: Causation

A special case of external conditioning arises when the conditioning event is a necessary antecedant to the target event. For example, let A be the event that S will die from cancer, and let X be the event that S will develop cancer at some time during his life. A formal way of expressing the special feature of this situation is to say $\pi(A|\bar{X}) = q(A|\bar{X}) = 0$. S will then provide $q' = q(A)$, $q_1 = q(A|X)$ and $q_3 = q(X)$, and coherence imposes the constraint $\pi(A) = \pi(A|X)\pi(X)$, or $\pi' = \pi_1\pi_3$ which is non-linear. As before, we can think of S as providing two estimates of π': q' directly, and q_1q_3 indirectly. If the former has variance σ^2, the latter, assuming all values stated by S are equally precise and uncorrelated, has approximate variance $(q_1^2 + q_3^2)\sigma^2$. Hence, weighting inversely proportional tc the variance, we have for a reconciled value, $\hat{\pi}'$, the result

$$(33) \qquad \hat{\pi}' - q' = \frac{q_1q_3 - q'}{1 + q_1^2 + q_3^2}$$

with precision $1 + (q_1^2 + q_3^2)^{-1}$ times the original precision.

Suppose S assesses the probability of his death from cancer as $q(A) = 0.35$, the probability of his acquiring a cancer as $q(X) = 0.40$, and finally the probability of dying from an acquired cancer as $q(A|X) = 0.70$. With $q' = 0.35$, $q_1 = 0.70$ and $q_3 = 0.40$, the reconciled value is 0.31; and the precision is increased 2.54 times its original value.

Notice that this method is especially valuable when q_1 and q_3, and hence q', are small, since then the increase in precision will be greatest. However, the probability metric may not be too appropriate for small probabilities and a log-odds metric preferred.

4.8 Internal Conditioning

External conditioning deals with the case where one extends the conversation to events not of immediate interest. If there is more than one target event, then conditioning may be used without bringing in additional events. For example, with two target events A and B, S may provide probabilities for all events in the induced partition, AB, $A\bar{B}$, $\bar{A}B$ and $\bar{A}\bar{B}$. Alternatively S may consider A, B conditional on A, and finally B conditional on $\bar{A}$. Of course the roles of A and B may be reversed. This structure is called internal conditioning since the conditioning events are among the target events. Another example arises with a partition $A_1, A_2, \ldots A_n$ of the type considered in Section 4.1. The probabilities $q(A_i)$ may be assessed directly, but another possibility is to consider the values exemplified by $q(A_1 | A_1 \cup A_2)$. The medical example (Table 4-1) is a case in point where S might consider the probability of dying from cancer conditional on his dying a natural death. There are numerous possibilities but we only explore in detail the above fourfold partition.

Suppose S is coherent in that his stated probabilities for a partition do add up to one, but they may exhibit more subtle types of incoherence. He can then assess the situation for two events A and B by first giving $q(AB)$, $q(A\bar{B})$ and $q(\bar{A}B)$--when $q(\bar{A}\bar{B})$ will be one minus the sum of these. He can then look at the situation conditionally using $q(A)$, $q(B|A)$ and $q(B|\bar{A})$. These two triples provide alternative complete specifications for S's opinions about the two events. The coherence constraints such as $\pi(AB) = \pi(A)\pi(B|A)$ --there are three of them--are non-linear, and a complete treatment must depend on numerical procedures: but an approximate argument is available. This utilizes the external conditioning argument of Section 4.7, treating any member of the partition, AB say, as having a necessary

antecedant, in this case A. Hence, the three assessments $q(AB)$, $q(A)$, and $q(B|A)$ can be taken together and reconciled to give, equation (33),

$$(34) \qquad \hat{\pi}(AB) - q(AB) = \frac{q(A)q(B|A) - q(AB)}{1 + q(A)^2 + q(B|A)^2} .$$

This can be used to provide reconciled values for all elements of the partition. Since they are weighted averages of values that do add to one (by assumption), they themselves will also add to one. A reconciled value $\hat{\pi}(A)$ may be obtained as $\hat{\pi}(AB) + \hat{\pi}(A\bar{B})$.

If S also conditions on B and provides $q(B)$, $q(A|B)$ and $q(A|\bar{B})$, then we may argue as follows: S has provided three assessments $q(AB)$ directly, $q(A)q(A|B)$ and $q(B)q(A|B)$ indirectly. These have variances σ^2, $[q(A)^2 + q(B|A)^2]\sigma^2$ and $[q(B)^2 + q(B|A)^2]\sigma^2$ respectively, the last two being approximate. They may be combined (Section 4.7) as a weighted average with weights inversely proportional to the variances, to give

$$(35) \qquad \hat{\pi}(AB) = \frac{q(AB) + \dfrac{q(A)q(B|A)}{q(A)^2+q(B|A)^2} + \dfrac{q(B)q(A|B)}{q(B)^2+q(A|B)^2}}{1 + [q(A)^2+q(B|A)^2]^{-1} + [q(B)^2+q(A|B)^2]^{-1}}$$

with precision given by the denominator times σ^{-2}.

To illustrate, consider the results of a forthcoming election in a particular district in the United States. Suppose there are two candidates (a Republican and a Democrat) for the Senate, and two candidates (a Republican and a Democrat) for the House. Let A and B, respectively, denote a democratic victory in the election for the Senate and the House. Suppose further that some political observer provided the assessments displayed in the upper part of Table

4-2. Thus $q(AB) = 0.1$, $q(B) = 0.4$, $q(A|\bar{B}) = 0.6$, etc. As might be expected, the assessments are inconsistent. The reconciled values, computed according to equation (34), using either A or B as the conditioning event, are displayed in the middle part of the table. The lower part of Table 4-2 presents the complete reconciliation based on all the available information using equation (35); it also presents the precisions of the reconciled values expressed as multiples of the precisions of the stated values. Thus, $\hat{\pi}(AB) = 0.129$ with a precision 7.9 times that of the original value 0.1. The reconciled value for $\pi(A)$ will be $0.129 + 0.340 = 0.469$. It is not possible to cite the precision of this value since the two values that have been combined are not independent; the sum $7.9 + 3.7 = 11.8$ is probably not too far wide of the mark.

ASSESSMENTS

	A	$\bar{A}$	Margin B	Conditional for A
B	0.1	0.4	0.4	0.3
$\bar{B}$	0.3	0.2		0.6
Margin A	0.5			
Conditional for B	0.3	0.5		

RECONCILIATION

Reconciliation using A as Conditioning Event		Reconciliation using B as Conditioning Event	
0.137	0.300	0.116	0.327
0.329	0.233	0.335	0.226

Complete Reconciliation		Precisions for Complete Reconciliation	
0.129	0.293	7.9	4.5
0.340	0.236	3.7	4.9

Table 4-2
Assessed and Reconciled Values for the Election Problem

5.0 DISCUSSION

In his treatise on the foundations of statistics, Savage writes:

> "According to the personalistic view, the role of the mathematical theory of probability is to enable the person using it to detect inconsistencies in his own real or envisaged behavior. It is also understood that, having detected an inconsistency, he will remove it. An inconsistency is typically removable in many different ways, among which the theory gives no guidance for choosing. Silence on this point does not seem altogether appropriate, so there may be room to improve the theory here..."
>
> "To approach the matter in a somewhat different way, there seem to be some probability relations about which we feel relatively "sure" as compared with others. When our opinions, as reflected in real or envisaged action, are inconsistent, we sacrifice the unsure opinion to the sure ones. The notion of "sure" and "unsure" introduced here is vague, and my complaint is precisely that neither the theory of personal probability, as it is developed in this book, nor any other device known to me renders the notion less vague." Savage (1954, pp. 57-58)

The present paper represents an attempt to deal with the issues raised by Savage, namely, the resolution of inconsistencies and the weighting of opinions. Our approach is based on a division of labor between a fallible subject S and a coherent investigator N who uses S's assessments to estimate S's "true" probabilities in the internal method,

or to update his own beliefs about the world in the external method. Naturally, such an approach encounters both philosophical and practical difficulties.

Perhaps the most obvious philosophical objection pertains to the coherence of N. Why permit first-order incoherence of q but exclude second-order incoherence of p? If people are inevitably fallible, it is reasonable to postulate a coherent N? Alternatively, one could argue, if S has an access to a coherent N, why permit inconsistency in the first place?

The coherence of N is needed to ensure the coherence of the reconciled values. If the core distributions, for example, are also allowed to be incoherent, they must be reconciled before they can be used to reconcile the basic assessments. This leads to an infinite regress that can be avoided only by assuming coherence somewhere in the process. Indeed, it does not appear unreasonable to assume a fallible assessor who is capable--in a more reflective mood and perhaps with the help of paper and pencil--of detecting and reconciling his own inconsistencies.

Notice that once N's coherent assessments, the core distributions, are determined, they are available for the resolution of any problem posed by S. In other words, the role of N is to link together the possible situations that S might face. This is in the true spirit of coherence in which a problem is not considered in isolation but viewed in conjunction with other problems, both real and conjectural. If the reader considers a simple example, he can easily conclude that our methods are unnecessarily involved--easier, he might say to do a naive reconciliation rather than determine the core distributions and then use these to effect the reconciliation. It is only when a set of examples

is considered that the power of the methods is revealed, for then, a simple determination of core elements serves for the whole set.

The present approach can be viewed as a compromise between two extreme positions on the nature of probability assessments: the rationalistic position that assumes coherence, and the empiricistic position that denies it. Neither position deals with the reconciliation problem; the former effectively ignores the issue, while the latter cannot solve it. By modeling a person as composed of a fallible S and a coherent N, we attempt to provide a more realistic idealization which could, nevertheless, be used to achieve rational reconciliation.

A central feature of the approach developed in this paper is the reliance on the core distributions and Bayes' rule to reconcile incoherence. Alternatively, one could start by considering the set of all reconciled values and then introduce criteria or axioms that restrict the choice of an admissible reconciliation. For example, if $q(A) = .62$ and $q(\bar{A}) = .34$, one may wish to restrict π so that $.62 \le \pi(A) \le 1-.34$. Additional constraints could further restrict the set of admissible values. This approach represents a viable alternative to the present procedure. It remains to be seen, however, whether one could develop a compelling set of criteria for reconciliation that would lead to a unique, or at least a highly constrained, solution.

From a practical standpoint, the major obstacle to the application of the proposed procedure is the difficulty in assessing the core distributions. All three distributions are readily interpretable: $p(A)$ is N's prior, $p(\pi|A)$ describes the relation between S's true beliefs and the external world, and $p(q|\pi)$ describes the relation between S's assessments and his true beliefs. Although these

expressions are psychologically meaningful, their assessment may prove very difficult in many cases. We are keenly aware of this problem, and much of the specific assumptions discussed in Sections 3.0 and 4.0 were introduced to simplify the assessment of the core distributions. It remains to be demonstrated that this information can be elicited from laymen and/or experts for non-trivial problems.

Reflection suggests to us that the introduction of these core distributions, and, in particular, the occurrence of precisions and correlations, is a reasonable, and perhaps necessary, requirement in the real-world situation before reconciliation is possible.

In the light of these difficulties, it could be argued that instead of applying the formal procedures developed in this paper, we could simply instruct the subject to resolve his own inconsistencies in the way that he finds most appropriate. Although this approach could often be employed, we believe that an explicit model provides a useful tool for the analysis of coherence. It focuses attention on the data needed to resolve incoherence, and it provides a rational procedure for reconciliation.

Of the three elements in the core distributions, $p(A)$ and $p(q|\pi)$ are of familiar types, but the third, $p(\pi|A)$ is rather novel. It provides, together with $p(\pi|\bar{A})$, an expression of S's ability as a probability assessor; and, in particular, a statement of S's variability. In effect, it looks at S as a diagnostic instrument: in medical terms, if the patient has appendicitis, A, what probability is the doctor, S, going to assign to A; and similarly for a patient with abdominal pain not originating from a ruptured appendix. We have seen that it is related to the calibration concept, $f(A|\pi)$. It leads, for example in equation (9), to unexpected

adaptations from π (or q) to revised probabilities for A, $p(A|\pi)$ or $p(A|q)$, which cannot always be π or q.

Simon French (1977) has suggested to us that $p(\pi|A)$ might depend on $p(A)$. In words, N's assessment of S might depend on his own views about A, the event under discussion. He cites the example where past experience has shown that in similar situations, N and S have tended to agree: then $p(\pi|A)$ might peak around $\pi = p(A)$. If this is admitted, we are lead to a curious probability, $p(\pi|A,p(A))$ and an unexpected form of Bayes Theorem that French has considered, namely

$$p(A|\pi,p) \propto p(\pi|A,p)p(A)$$

where $p = p(A)$. This is an interesting idea that introduces a reasonable correlation between N and S, but it should be noted that our model partly allows for such correlation. We suppose π is unaffected by p, given A and given $\bar{A}$, but this will not imply that π is unconditionally free of p; indeed, quite the contrary if S is a good appraiser.

There are at least two ways in which the model can be generalized. First, the world external to both N and S can contain uncertain quantities and not just uncertain events. This has been discussed by Morris (1974, 1977). An example might be the case where the meteorologist is forecasting the amount of rain tomorrow rather than whether or not it will rain.

A second generalization is to the case where the data-base for S changes. In our discussion it has been supposed fixed. An example arises when the meteorologist, forecasting the weather on day 2 learns about what the weather has been on day 1. The role of the data-base certainly needs more

examination: it might resolve the dilemma proposed by French, in that the data common to N and S might explain the possible correlation suggested by him.

REFERENCES

Brown, R. V. and Lindley, D. V. Notes on rationality and the reconciliation of incoherence. In Watson, S. R., Brown, R. V. and Lindley, D.V. Three Papers on the Valuation of Decision Analysis. Technical Report 77-2. McLean, VA: Decisions & Designs, Inc., 1977.

de Finetti, B. La prevision: ses lois logiques, ses sources subjectives. Ann. Inst. H. Poincare, 1937, 7, 1-68. Translated into English in H. E. Kyburg, Jr., and H. E. Smokler (Eds.), Studies in subjective probability. New York: Wiley, 1964. Pp. 93-158.

French, S. The use of Bayes' theorem to update belief when the likelihood depends on the prior. Department of Decision Theory, University of Manchester, November 1977.

Kahneman, D. and Tversky, A. Intuitive prediction: Biases and corrective procedures. Management Science, 1978, in press.

Krantz, D. H., Luce, R. D., Suppes, P., and Tversky, A. Foundations of measurement (Vol. 1). New York: Academic Press, 1971.

Morris, P. Decision and analysis expert use. Management Science, 1974, 20, 1233-1241.

Ramsey, F. P. Truth and probability. In F. P. Ramsey, The foundations of mathematics and other logical essays. New York: Harcourt, Brace, 1931. Pp. 156-198. Reprinted in H. E. Kyburg, Jr., and H. E. Smokler (Eds.) Studies in subjective probability. New York: Wiley, 1954.

Savage, L. J. The foundations of statistics. New York: Wiley, 1954.

Slovic, P., Fischhoff, B. and Lichtenstein, S. Behavioral decision theory. Annual Review of Psychology, 1977, 28, 1-39.

Spetzler, C. S. and Stael von Holstein, C. A. S. Probability encoding in decision analysis. Management Science, 1975, 22, 340-358.

Tversky, A. and Kahneman, D. Judgment under uncertainty: Heuristics and biases. Science, 1974, 135, 1124-1131.

Winkler, R. Scoring rules and the evaluation of probability assessors. Journal of the American Statistical Association, 1969, 64, 1073-1078.

CONTRACT DISTRIBUTION LIST
(Unclassified Technical Reports)

Director, Engineering Psychology Programs, Code 455 Office of Naval Research 800 North Quincy Street Arlington, VA 22217	5 copies
Commanding Officer ONR Branch Office Attention: Dr. J. Lester 495 Summer Street Boston, MA 02210	1 copy
Commanding Officer ONR Branch Office Attention: Dr. Charles Davis 536 South Clark Street Chicago, IL 60605	1 copy
Commanding Officer ONR Branch Office Attention: Dr. E. Gloye 1030 East Green Street Pasadena, CA 91106	1 copy
Office of Naval Rsearch International Programs Code 102IP 800 North Quincy Street Arlington, VA 22217	6 copies
Director, Naval Research Laboratory Technical Information Division Code 2627 Washington, DC 20375	6 copies
Defense Documentation Center Cameron Station Alexandria, VA 22314	2 copies

Supplemental Distribution List

Director, Naval Analysis Programs
Code 431
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Director, Operations Research
Programs, Code 434
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Director, Statistics and
Probability Program, Code 436
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Dr. Bruce McDonald
Office of Naval Research
Scientific Liaison Group
American Embassy, Room A-407
APO San Francisco 96503

Navy Personnel Research and
Development Center
Management Support Department
Code 210
San Diego, CA 92152

Dr. Edgar M. Johnson
Organization and Systems
Research Laboratory
U.S. Army Research Lab
5001 Eisenhower Avenue
Alexandria, VA 22333

Air Force Office of Scientific
Research
Sciences Directorate, NL
Bolling Air Force Base
Washington, DC 20332

Dr. Paul Slovic
Decision Research
1201 Oak Street
Eugene, OR 97401

Dr. Ron Howard
Stanford University
Stanford, CA 94305

Dr. Ward Edwards
Director, Social Science Research
Institute
University of Southern California
Los Angeles, CA 90007

Dr. Charles Gettys
University of Oklahoma
Department of Psychology
455 West Lindsey
Norman, OK 73069

Dr. Gary McClelland
University of Colorado
Institute of Behavioral Sciences
Boulder, CO 80309